IDENTIFYING STUDENT- AND CLASS-LEVEL CORRELATES OF SIXTH-GRADE STUDENTS' LISTENING COMPREHENSION

HELEEN BOURDEAUD'HUI*, KOEN AESAERT**, & JOHAN VAN BRAAK*

** Ghent University, ** Leuven University*

Abstract
Despite the importance of listening, little investigation of potential correlates of listening comprehension in the language of schooling is done. Therefore, the purpose of this study was to investigate which student- and class-level characteristics are related to sixth-grade students' listening skills in Flanders. A sample of 974 students in 70 classes completed a listening test in order to gather information on their ability to understand and interpret oral information. Further, different questionnaires were administered to the students, their parents and teachers. The results of the hierarchical regression analysis with multilevel design showed that the differences in listening comprehension skills could be primarily attributed to differences in student-level characteristics. The results indicated that students with higher working memory ability, more vocabulary knowledge and lower extrinsic listening motivation performed significantly better on the listening test. In addition, the educational level of the parents and the language diversity in the class was significantly related to students' listening skills in the language of schooling. This study is an important starting point in unraveling the black box of listening skills in the elementary school context. Suggestions for further research and practice were made.

Keywords: listening comprehension, language of schooling, elementary school, multilevel analysis, student- and class-level correlates.

Bourdeaud'hui, H., Aesaert, K., Van Braak, J. (2020). Identifying student-and class-level correlates of sixth-grade students' listening comprehension. L1-Educational Studies in Language and Literature, 20, 1-38. https://doi.org/10.17239/L1ESLL-2020.20.01.04
Corresponding author: Heleen Bourdeaud'hui, Department of Educational Studies, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium. email: Heleen.Bourdeaudhui@ugent.be.

1. INTRODUCTION

In the elementary class context, listening comprehension is an important skill for everyday functioning. The traditional teaching approach is still heavily dependent on listening to the teacher, and students spend more time listening as a way to learn than they do using other communication skills (Hogan, Adlof, & Alonzo, 2014). For example, students need effective listening skills to understand orally presented information or instructions from the teacher. In addition, listening comprehension has a significant influence on the development of students' reading comprehension (Adlof, Catts, & Little, 2006). For these reasons, it is no surprise that the ability to listen in the language of schooling (further referred to as Lsch) is regarded as a precondition for effective learning (Adelmann, 2012; Hogan et al., 2014; Wolfgramm, Suter, & Göksel, 2016). Despite its importance, most elementary school teachers mainly focus their attention on reading and writing instruction, while listening (and speaking) has less common been taught in the class (Beall, Gill-Rosier, Tate, & Matten, 2008; Siegel, 2014). Likewise, listening in the Lsch has rarely been studied in depth in effectiveness research (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012), except for the substantial body of work investigating the role of working memory in listening skills (e.g., Kim, 2016; Tighe, Spencer, & Schatschneider, 2015). This lack of literature may be due to prevailing perceptions towards listening and listening education. For a long time, there was the implicit assumption that listening skills in the Lsch are already developed before entering elementary school. However, more recent research shows that it is not the case that all students perform naturally well on listening tests, and listening skills might be enhanced through instruction (Brown, 2008; Fogelsong, 2016; Lau, 2017). In Flanders, i.e., the Dutch speaking part of Belgium, about 10% of the native Dutch-speaking and 40% of the non-native Dutch-speaking students do not achieve the standards for listening in Dutch at the end of elementary school (Denis, Dierick, Janssen, & Aesaert, 2019). To improve students' listening skills, it is important to identify the different characteristics that affect listening and to organize the learning environment accordingly.

Besides, existing listening research appears to have been mostly conducted from a single-level student perspective (e.g., Alonzo, Yeomans-Maldonado, Murphy, & Bevens, 2016; Lau, 2017; Marx & Roick, 2012; Wolfgramm et al., 2016). However, educational effectiveness research should be multilevel in nature as students are nested in classes and classes are nested in schools (Creemers & Kyriakides, 2010). Correlates operate at different levels and are expected to interact across levels, and thus ignoring variation between these levels may lead to an incomplete interpretation of listening skills. As such, the aim of the present study is to look in greater detail for both student- and class-level determinants that could explain listening differences across late elementary school students. An in-depth understanding of the characteristics, which contribute to variation in students' listening achievement, is essential to gain insight into the actual construct of listening comprehension, to

increase our understanding of why some students are more successful listeners than others, and to improve listening education quality.

2. LITERATURE REVIEW

2.1. Listening comprehension skills

Listening comprehension skills refer to the ability to process, integrate, and understand the meaning of what has been said (Hogan et al., 2014). Listening is one of the four basic language skills, next to reading, writing and speaking. Listening and reading require students' ability to decode (receptive skills), whereas speaking and writing refer to students' skills to express themselves (productive skills) (Brownell, 2012). Although listening and reading require different decoding processes (i.e., visual versus oral input), language researchers indicate that they share common cognitive activities during information processing (Anderson, 2000). Both listening and reading require the interaction of bottom-up and top-down processes (Field, 2004). Students draw on bottom-up processing when they use their linguistic knowledge to interpret small units such as words or sentences for composing the auditory message. Students rather use top-down processing when they rely on contextual or prior knowledge to build a conceptual framework and understand the message as a whole. During listening, these two processes typically work as an "interactive-compensatory" mechanism, under the limitations of the capacity of the working memory (Field, 2004; Lau, 2017; Vandergrift, 2004). Further, as is the case with reading comprehension, researchers generally distinguish between literal and inferential listening comprehension skills (Kim, 2016; Ulu, 2016). Literal listening refers to the understanding of the explicit meaning and mainly requires the accurate storage of information, such as recalling details or facts from a text passage. On the other hand, inferential listening refers to the comprehension of the implicit message the speaker is trying to give. Inferential comprehension tasks require the processing of information, such as making inferences between different text parts and maintaining the coherence of the text (Kim, 2016; Ulu, 2016). In this study, both literal and inferential listening skills will be measured through a listening comprehension test. In the following sections, we describe in more detail research on student- and class-level correlates of listening comprehension in the Lsch.

2.2. Identifying student-level correlates of listening skills

The student-level explanatory variables that have been mostly studied in relation to listening comprehension in the Lsch include cognitive and linguistic variables, such as working memory and vocabulary knowledge. Other student-level factors which are often linked to language skills, such as the influence of strategy use or motivation (e.g., De Smedt, Van Keer, & Merchie, 2016) are less frequently studied in research on listening comprehension in the Lsch. Below, we categorize student-level

explanatory variables of listening skills into cognitive, linguistic, motivational, and background characteristics.

Cognitive student characteristics. *Working memory,* which refers to the ability to actively keep task-relevant information in mind while processing information (Just & Carpenter, 1992), has proven to be an important predictor of listening comprehension in different studies in elementary and secondary grades (Alonzo et al., 2016; Florit, Roch, Altoè, & Levorato, 2009; Kim, 2016; Molloy, 1997; Tighe et al., 2015). Working memory is expected to be of high relevance for students' listening performance, as during a listening task, students have to temporarily store linguistic input, while they process and integrate it with new incoming information (Florit et al., 2009; Kim, 2016; Tighe et al., 2015). In this way, working memory is a foundational skill for creating local (within a sentence) and global (across sentences) coherence in the oral text (Kim, 2016). Although both reading and listening apply on working memory, listening skills require a greater processing load, because the speaker sets the processing rate and the text does not remain available after listening (Roch, Florit, & Levorato, 2012). Further, working memory may play a more important role for non-native speaking students, because their lower knowledge of vocabulary in the Lsch could cause an excessive load of the working memory (Wolfgramm et al., 2016). A second cognitive characteristic, which has often been linked to language success, is students' *strategy use*. Listening strategies can be described as listeners' intended plans and mental operations to handle and comprehend incoming speech, such as elaborating or making inferences (Field, 2010; Lau, 2017). Different researchers studying listening in learning a second language have highlighted the importance of strategy use for successful listening, showing that good and poor listeners differ in the frequency and the quality of their strategy use (e.g., Berne, 2004; Graham, 2017). However, research investigating the influence of listening strategy use on listening skills in the Lsch, is barely available. Only Lau (2017) provided empirical evidence for strategy use as a crucial factor in listening in the Lsch, showing that high proficiency Chinese listeners possess more types of cognitive strategies and use them more frequently and effectively than low proficiency listeners. As such, it can be expected that more listening strategy use will be positively correlated with higher listening outcomes.

Linguistic student characteristics. Different studies have demonstrated *vocabulary knowledge* to be of particular importance for elementary school students' listening skills (e.g., Alonzo et al., 2016; Andringa et al., 2012; Wolfgramm et al., 2016). Especially children's receptive word knowledge and their word recognition are necessary to process the meanings of words from the input text (Kim, 2016; Tighe et al., 2015). If listeners have not recognized enough words in the spoken text, they will not be able to construct an adequate representation of the text (Staehr, 2009), and their lexical gaps will interrupt the continuous process of listening due to the transient nature of spoken texts (Hagtvet, 2003).

Motivational student characteristics. The listening process is not only influenced by cognitive and linguistic skills, but motivational processes can be involved as well.

The intrinsic and extrinsic orientations of motivation, as conceptualized in the self-determination theory, constitute a useful framework for studying motivation in educational contexts (Ryan & Deci, 2002). *Intrinsic motivation* refers to internal factors such as enjoyment and internal satisfaction for learning (Stutz, Schaffner & Schiefele, 2017). In this way, intrinsically motivated students may listen because they are interested in listening activities and listen for pleasure and enjoyment. *Extrinsic motivation* is based on attaining instrumental aims that lie beyond the actual learning process (Stutz et al., 2017). As such, extrinsically motivated students may listen in order to get praise by their parents or the teacher. In Flanders, national assessment research showed that intrinsic reading motivation was positively correlated with sixth-grade students' reading and listening skills. In contrast, extrinsic reading motivation was negatively correlated with both reading and listening skills (Denis, Dierick, Janssen, & Aesaert, 2019). Furthermore, research investigating the link between different listening motivation orientations and listening skills in the Lsch is limited. Listening motivation in learning a second language has been studied more extensively, but the results were rather inconsistent. Some studies found a positive correlation between L2 listening comprehension and intrinsic listening motivation (e.g., Baleghizadeh & Rahimi, 2011), although other studies could not find a significant correlation (e.g., Harputlu & Ceylan, 2014; Vandergrift, 2005). Considering extrinsic listening motivation, most studies could not find a significant relationship with L2 listening comprehension (e.g., Baleghizadeh & Rahimi, 2011; Vandergrift, 2005).

Student background characteristics. Next to the above-mentioned cognitive, linguistic, and motivational characteristics, it is essential to take into account individual background characteristics in explaining variation in listening comprehension. Different student background characteristics, such as gender, home language, having a learning disorder, grade retention, parents' educational level, and educational support at home have frequently proven to influence students' language skills and must be integrated as control variables in the statistical model. As for the role of *gender*, there is a body of research showing that elementary school girls have a clear advantage in language skills, such as reading (Logan & Johnston, 2009) or writing (De Smedt et al., 2016). However, findings on gender effects in listening comprehension are rather unclear. Although some studies have found that elementary school girls outperform boys in listening (e.g., Oduolowu & Oluwakemi, 2014), other studies could not find a significant relationship (e.g., Lehto & Antilla, 2003; Lin, Liu, Chen, Wang, & Kao, 2015), or even found a gender difference in favor of boys (Wolfgramm et al., 2016). In addition, various researchers indicate that *non-native* speaking students experience a greater challenge for listening skills in the Lsch in comparison with native speaking students (Andringa et al., 2012; Marx & Roick, 2012; McKendry & Murphy, 2010; Wolfgramm et al., 2016). It makes sense that non-native speaking students have more constrained vocabulary knowledge in the Lsch than native speaking students (Staehr, 2009; Marx & Roick, 2012), but research also showed that both groups do not rely on the same processes in listening comprehension. Non-native speaking students make more use of lexical information or bottom-up

processing and are less able to use syntactic cues or top-down processing in monitoring tasks (Andringa et al., 2012). In addition, they are less effective in relating prosodic cues to semantic information (Akker & Cutler, 2003). Further, children with a *learning difficulty*, such as dyslexia or attention problems, can be disadvantaged for listening skills because they often suffer from difficulties with lower working memory, more limited vocabulary, and certain aspects of comprehension, such as making inferences (Jeffries & Everatt, 2004; McInnes, Humphries, Hogg-Johnson, & Tannock, 2003). With regard to *grade retention*, research indicates that retained students are often disadvantaged, as the learning gain for language development has already disappeared at the end of the next grade (Huddleston, 2014; Vandecandelaere, Vansteelandt, De Fraine, & Van Damme, 2016). Finally, indicators of family background such as the *educational level of the parents* have been directly associated with language skills and must be included as control variables in the model. For example, Acat, Demiral and Kaya (2016) found that elementary school students' listening skills are positively correlated with the educational level of the mother and the father. In this regard, *a literacy-supportive home environment*, where parents offer a large amount of learning resources and often spend learning time with their children, has consistently been found to be a significant predictor of students' language outcomes (Klauda, 2009; Mullis, Martin, Foy, & Hooper, 2017).

2.3. Identifying class-level correlates of listening performance

Although much of the variance in student achievement can be attributed to student-level characteristics, differences at the class level may also significantly influence student outcomes. Below, we categorize class-level explanatory variables of listening skills into teacher practices, attitudes, expectations, and background characteristics.

Teacher practices. Effective listening skills are crucial to school success and a large amount of class time is spent listening. However, the knowledge about what teachers actually do regarding listening instruction is very scarce due to the lack of listening research studying the practices that teachers employ in the classes (Beall et al., 2008; Fogelsong, 2016). This neglect may be created by the expectations that students are already prepared to listen effectively when they come to elementary school (Lau, 2016; Siegel, 2014). In addition, listening has often been overlooked in teacher preparation programs and most coursework do not provide explicit listening instruction practices for teachers (Clark, Scruggs, & Szydlowski, 1999; Graham, Santos, & Vanderplank, 2011). However, in the past 20 years, studies about listening in educational settings indicate that instruction can enhance student's listening achievement (Brownell, 2012; Cohen & Wolvin, 2011; Imhof, 2008; Jalongo, 2008), providing a strong rationale for further investigation.

Teacher attitudes. Whereas teacher practices influence learning directly, teacher attitudes are theorized to influence student learning more indirectly through their association with instructional practices (Palardy & Rumberger, 2008). The attitude of the teacher towards listening and listening instruction may be an important element

in listening achievement, because it can affect how teachers behave in the classroom through instructional strategies and interactions with students. We expect that teachers with positive attitudes related to listening will spend more time on listening instruction and more frequently integrate listening in their lessons, resulting in higher listening outcomes. There is a need to understand how teachers presently interpret listening in their classrooms and how this understanding informs their choices for instruction (Siegel, 2014).

Teacher expectations. Finally, among class-level factors, the expectations of the teachers towards their students, i.e., their perception of student achievement, may influence students' listening skills. Teachers with higher expectations will more positively interact with their students and enhance student learning, whereas teachers with lower expectations may think their students lack the necessary traits for success and have a detrimental effect (Hassel & Ridout, 2017; Van den Bergh, Denessen, Hornstra, Voeten, & Holland, 2010).

Class background characteristics. Next to the above-mentioned teacher attitudes and teacher practices, different teacher background factors, such as gender, certification, years of experience, class size, and class composition can be related to students' listening skills and must be taken into account as control variables in the statistical model. Considering *teacher certification*, prospective teachers studying at bachelor level are more generally informed on how to teach language skills. At master level, student teachers receive further scientific background and research studies of language courses, possibly leading to more tools and a deeper understanding of instructional practices (Akbari & Dadvand, 2011). Further, a consensual finding is that *teacher experience* has a significant positive effect on students' language achievement, at least during the first five years in the class (Croninger, Rice, Rathbun, & Nishio, 2007). With regard to *class size*, there is a growing consensus that students in smaller classes reach higher levels of language skills than students in larger classes (Fredriksson, Ockert, & Oosterbeek, 2013). Students in smaller classes may develop better listening skills as listening is easier when students can make direct eye contact with and are in close physical proximity to the speaker, whereas students in larger classes are more likely to engage in passive listening behavior (Blatchford, Bassett, & Brown, 2011). In addition, in a larger class, one-to-one interaction between students and their teacher might be more difficult than in smaller classes. Finally, research on group composition effects should also be considered, especially since statistically negative effects of more *language diversity* were found (Belfi, Haelermans, & de Fraine, 2016; Kyriakides, Creemers, & Charalambous, 2019). The language contact hypothesis of Driessen (2003) states that students in segregated classes have fewer opportunities to come into contact with the language of instruction when compared to students in classes with a higher proportion of native-speaking students.

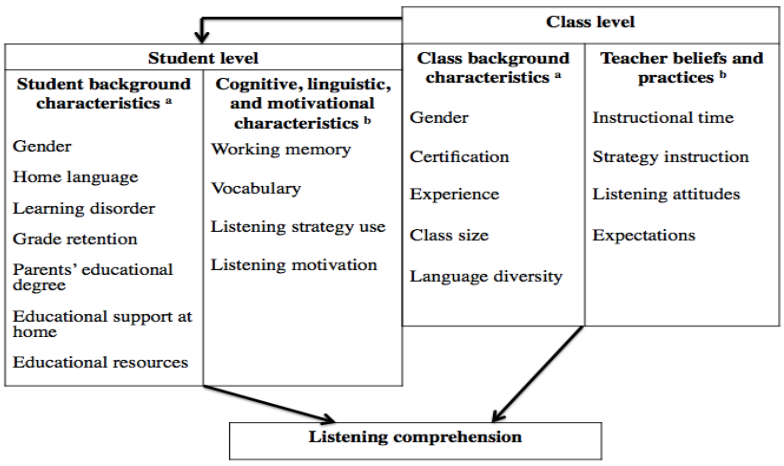
3. OBJECTIVES

Listening skills are seen as an important factor in the elementary school context. However, listening skills in the Lsch are seldom studied in effectiveness research (Andringa et al., 2012; McKendry & Murphy, 2010). The current study aims to close a gap in the literature by investigating which student- and class-level characteristics are related to elementary school students' listening skills in the Lsch, after controlling for different background characteristics. Figure 1 shows which characteristics are integrated as control and explanatory variables in the multilevel model based on the literature study. We specifically focus on upper-elementary grades (i.e., grade six) because these students can engage in higher-order listening skills, while students in lower grades are still developing lower-order listening skills. While taking the above-mentioned theoretical remarks into account, our research hypotheses are:

(1) Cognitive, linguistic, and motivational student-level characteristics, i.e., higher working memory, listening strategy use, vocabulary knowledge, and intrinsic listening motivation are positively related to sixth-grade students' listening skills in the Lsch.

(2) Beliefs and practices of the teacher, i.e., higher instructional listening time and instruction in listening strategies, listening attitudes, and teacher expectations are positively related to sixth-grade students' listening skills in the Lsch.

Figure 1. Overview of the multilevel model (a = Control variables, b = Explanatory variables)



4. METHODOLOGY

4.1. Sample

Different tests and questionnaires were administered to a sample of 974 sixth-grade elementary-school students in Flanders, the Dutch-speaking part of Belgium. The data were collected in 51 schools and 70 classes from March till May 2018, with a mean of 19 students per class. The selected schools are a representative sample of Flemish schools as they were stratified for school size (i.e., small school < 180 students; large school ≥ 180 students), province, and educational network (i.e., official public education, subsidized public-authority education, and subsidized private-authority education). Of the 70 elementary school teachers, 75.7% were female and 24.3% were male. The teaching experience of the teachers was on average 14 years, with a minimum of 0 and a maximum of 38 years of experience. The mean age of the students was 11.88 years, with a minimum of 10.57 and a maximum of 14.54 years. The present study involved both native (*n* = 812) and non-native (*n* = 162) Dutch-speaking students. Besides Dutch, fifteen different home languages were represented in the study, i.e., students were speaking French (*n* = 69), Arabic (*n* = 21), Turkish (*n* = 17), Berbers (*n* = 11), Polish (*n* = 11), English (*n* = 10), Russian (*n* = 5), Greek (*n* = 5), Spanish (*n* = 3), Portuguese (*n* = 3), Japanese (*n* = 2), German (*n* = 2), Bulgarian (*n* = 2), and Italian (*n* = 1) as their first language. Of the 974 students, 50.6% (*n* = 493) were boys and 49.4% (*n* = 481) were girls. Further, 131 students were diagnosed with a developmental disorder. More specifically, 48 students were diagnosed with dyslexia, 24 with dyscalculia, 34 with an attention deficit disorder, 22 students showed autism spectrum symptoms, and 3 students had a hearing problem. All students completed a comprehensive listening test, a working memory test, a vocabulary test, and a background questionnaire. Each student was tested in two whole-class testing sessions, each lasting up to 50 minutes, separated by one to two weeks. At the first test moment, students had to fill in the first part of the listening test, the working memory test, and the student questionnaire. At the second test moment, students completed the second part of the listening test and the vocabulary test. Active informed consents were obtained from the parents of the students to use the data for further scientific research. Next to the student instruments, a questionnaire was administered to the teachers and the parents of the students.

4.2. Measurement instruments

Dependent variable: listening comprehension skills. Students' listening skills were measured through a validated and standardized Dutch audiovisual listening test (Bourdeaud'hui, Aesaert, & van Braak, submitted). The listening instrument was psychometrically validated through a large group of 1019 sixth-grade elementary-school students in Flanders. Item Response Theory (IRT) was used to investigate the item

characteristics and to provide evidence of the validity and reliability of the developed instrument. The listening test had an acceptable overall reliability (Expected A Posteriori = .67). A detailed description of the validation process of the listening test can be found in Bourdeaud'hui et al. (submitted). The listening test was developed based on different design principles. First, the test was offered in the format of a videotext as in most authentic listening situations the listener is able to simultaneously hear and see the speaker (Suvorov, 2015). Second, to prevent students' listening skills being influenced by their reading comprehension, every question and answering option has been read aloud on the recording (Chang & Read, 2013). Finally, in order to replicate real-world listening situations, every recording was played only once to the students, and note-taking was not allowed during the listening test (Green, 2017).

The listening test can be subdivided in two parts. First, students were asked to listen and watch to six informative video fragments, selected from a daily Dutch youth news program "Karrewiet". Second, students had to listen to two instructions with practical assignments from a fictional teacher—recorded by a professional, native Dutch female speaker. The listening test consisted of 24 multiple-choice and open-ended items, which were all scored dichotomously (1 = correct; 0 = incorrect), with a scoring key for the assessment of the open-ended items. The different test items assessed two listening levels: literal and inferential comprehension. More information about the listening test and some example items can be found in Appendix A.

4.3. Student-level variables

To measure students' *working memory* skills, a modified and Dutch version of the Listening span task of Molloy (1997) was used. Two modifications were made to the original instrument: (1) the sentences were translated from English to Dutch, and (2) the one-to-one test situation between the test taker and the student was adapted to whole class-assessment. For this reason, the working memory test was recorded on an audiotape, and the students had to write down their answers instead of giving a verbal response. During test administration, students had to listen to groups of sentences that could be sense or nonsense. As the task proceeded, the test increased in complexity, ranging from two sentences up to four sentences per group. After each group, students had to respond with true (sense) or false (nonsense) and write down the last word at the end of each sentence. For instance, students listened to the following group: "The dog slept in the house" and "The elephant read the sign". The students had to indicate respectively "true" and "false" and recall the final words: "house" and "sign". The adapted version of Molloy's working memory test was validated through Item Response Theory. The working memory test had a good internal consistency (EAP = .88). More information about the validation procedure of the working memory test can be found in Appendix B.

Listening strategy use was measured by means of a scale asking for students' self-reported use of listening strategies before, during, and after listening to the text

(e.g., "Before I start to listen, I have a plan in my head for how I am going to listen"). The scale was developed through selecting different strategies from listening strategy literature (e.g., Berne, 2004; Vandergrift, 2007) and was verified by means of exploratory and confirmatory factor analysis. Cronbach's alpha was .83, indicating good internal consistency of the scale. All eleven items were rated on a six-point Likert-type scale (1 = never, 2 = rarely, 3 = sometimes, 4 = regularly, 5 = mostly, 6 = always). For each student, a "listening strategy" score was calculated by dividing the mean score of the different items through 6 and multiplying by 100. More information about the general validation procedure of the different scales can be found in Appendix C. The items of the listening strategy scale can be found in Table D1 in Appendix D.

The data about students' *vocabulary knowledge* were gathered by means of a revised paper version of the Dutch online youth-word test (Brysbaert, Keuleers, Mandera, & Stevens, 2014). The test was adapted for sixth-grade students through the author of the online youth-word test. During this test, students were shown a random list of 20 existing or non-existing words and were asked to indicate which words exist. The non-words of the vocabulary test were constructed with the pseudoword generator Wuggy (Keuleers & Brysbaert, 2010). The vocabulary test was also validated through Item Response Theory and had a good overall reliability (EAP = .71). More information about the validation procedure of the vocabulary test can be found in Appendix B.

In order to measure students' *listening motivation*, the Reading Motivation Questionnaire for Elementary students (RMQ-E) (Stutz et al., 2017) was translated in Dutch and adapted to the context of listening (i.e., by replacing "reading" with "listening" to reflect the context of listening education). The questionnaire captures two factors: intrinsic motivation (e.g., "I listen because the things that are told are thrilling me") and extrinsic motivation (e.g., "I listen because I want to outperform others in the class"). The two-factor model of intrinsic and extrinsic motivation was verified by means of confirmatory factor analysis. Cronbach's alpha for the intrinsic and extrinsic listening motivation scale was respectively .85 and .68. All twelve items were rated on a six-point Likert-type scale (1 = totally disagree, 2 = disagree, 3 = rather disagree, 4 = rather agree, 5 = agree, 6 = totally agree). Intrinsic and extrinsic motivation scores were calculated by dividing the mean score of the different items through 6 and multiplying by 100. More information about the listening motivation items and the validation of the scale can be found in Table D2 in Appendix D.

Background data about the students were gathered using a student and parent questionnaire. Different student background variables (i.e., gender, home language, learning disorder, grade retention, education level of the parents, educational support at home, and educational resources at home) were included. *Gender* was dummy coded zero for boys and one for girls, while the most spoken language with the mother represented *home language* (0 = Dutch, 1 = other language). *Grade retention* was measured through a single item, which asked if students have already been retained in school (0 = no grade retention, 1 = grade retention). *Learning*

disorder was measured through questions asking for different learning or developmental disorders (0 = no, 1 = yes). In this study, the *educational level of the parents* was used as a proxy for socio-economic status (SES), as this is of major importance in Flanders and correlates highly with the occupational level and parental income (Van Laere, Aesaert, & van Braak, 2014). Educational level of the mother and the father was delineated on a four-point scale: 1 = elementary education, 2 = lower secondary education, 3 = higher secondary education, and 4 = higher education. *Home educational resources* referred to the availability of the following educational resources at home: computer, desk to study, quiet place to study, internet, dictionary, television, radio (choices: yes or no), and number of books at home (choices: very few, one bookshelf, one bookcase, two bookcases, three or more bookcases). *Educational support at home* was measured using nine items rated on a six-point scale (1 = never, 2 = rarely, 3 = sometimes, 4 = regularly, 5 = mostly, and 6 = always), which captures, for example, how often parents offer support with homework. This scale was verified by means of exploratory and confirmatory factor analysis. Cronbach's alpha was .74, indicating good internal consistency of the scale. More information about the items and the validation of this scale can be found in Table D3 in Appendix D.

4.4. Class-level variables

Teacher practices. To determine *the amount of time* spent in the class on instruction and evaluation in listening, two items had to be answered on an 8-point rating scale ranging (1 = about once a year, 2 = multiple times a year, 3 = about once a month, 4 = two or three times a month, 5 = about once a week, 6 = multiple times a week, 7 = about every day, and 8 = multiple times a day). Further, it was measured how often teachers spent time on instructing different *listening strategies* (e.g., summarizing the listening text). Eight listening strategy instruction items had to be answered on a six-point Likert-type scale (1 = never, 2 = rarely, 3 = sometimes, 4 = regularly, 5 = mostly, and 6 = always). This scale was verified by means of confirmatory factor analysis (see Table E1 in Appendix E). Cronbach's alpha was .76, indicating good internal consistency of the scale.

The *attitude of the teacher* towards listening was measured by means of ten items on a six-point Likert-type scale (1 = totally disagree, 2 = disagree, 3 = rather disagree, 4 = rather agree, 5 = agree, and 6 = totally agree). The scale measured teacher attitudes towards listening and listening education (e.g., "I think spending time on listening education is as important as spending time on other language skills") and was verified by means of confirmatory factor analysis (see Table E2 in Appendix E). For each teacher, the score of the different items was calculated by dividing the mean score of the different items through 6 and multiplying by 100. Cronbach's alpha for the teacher attitude scale was .76, indicating good internal consistency of the scale.

Data about the *expectations of the teacher* were gathered by using the teacher expectation scale of Thys and Van Houtte (2016). This scale consists of three items measuring the expectations that teachers held concerning their students' future school progress. The three items were: "I expect most students to perform well in their future school career", "I expect most students to perform well in secondary education", and "I think most students will find their way in work-life". Cronbach's alpha for the teacher expectation scale was .86, indicating good internal consistency of the scale. Appendix F gives an overview of all included independent variables.

Data about *class background characteristics* were gathered through a teacher questionnaire. Teacher background data encompassed variables related to teacher gender (0 = female, 1 = male), years of experience, and teachers' certification. *Teacher experience* was captured by the open question: By the end of this school year, how many years will you have been teaching in elementary school? *Teachers' certification* was measured on a qualification scale (1 = certified elementary school teacher, 2 = certified secondary school teacher, 3 = master's degree, 4 = master's degree in teacher education). *Language diversity* was determined by using the Herfindahl index. This index represents the variety of the languages spoken in the class and is calculated with the following formula: 1 - [(proportion of Dutch-speaking students)^2 + (proportion French)^2 + (proportion English)^2 + (proportion Turkish)^2 + (proportion Moroccans/Berbers)^2 + (proportion other language)^2] (Dronkers & van der Velden, 2013). A value of zero means that there is no language diversity in the class, with all students speaking the same language, whereas a value approaching one indicated high language diversity. *Class size* was operationalized as the number of students in a class. *Class resources* referred to learning materials that are allocated to the class, i.e., computer, radio, television, Internet, and Whiteboard.

4.5. Data analysis

The data were analyzed by means of a regression analysis with a multilevel design, which take into account the hierarchical structure of the data (Raudenbush & Bryk, 2002), that is, 974 students (level 1) were nested within 70 classes (level 2). Student- and class-level characteristics were entered stepwise in five subsequent models. In every step of the analysis, the significance of each parameter was tested and the model fit was evaluated with the statistical software program MLwiN 3.02 (Rasbash, Browne, & Charlton, 2018). In a first step, a fully unconditional null model without any predictor variables was tested (Model 0) to examine whether a multilevel approach was justified compared to a single-level regression analysis. In a second step, student background variables including gender, home language, learning disorder, grade retention, educational level of the parents, educational support at home, and educational resources at home were integrated as control variables (Model 1). Third, cognitive, linguistic, and motivational student characteristics, including working memory, vocabulary, strategy use, and intrinsic and extrinsic motivation were added to the model (Model 2). Fourth, background characteristics at the class level,

including teacher gender, years of experience, certification, language diversity, class size, and class resources were added (Model 3). Finally, teacher process characteristics, including practices, attitudes, and expectations were integrated (Model 4). By using a stepwise multilevel approach, the additional value of each subset of variables to the model was checked. Before each subset of variables was added to the model, non-significant factors were omitted to continue the analysis with a model with only significant factors. The change in deviance between models was used to investigate model improvement (Anderson, 2012).

5. RESULTS

5.1. Descriptive statistics

Descriptive statistics for sixth-grade students' scores on the listening test and the included variables were computed. Table 1 gives an overview of the mean scores, the standard deviations, and the minimum/maximum score for the variables. The students' level of listening comprehension is ranging from -2.96 to 2.97, with an average of .06 on the standardized listening test. Further, the results showed that teachers spend on average two or three times a month time on instructing listening skills, and about once a month time on evaluating listening skills. Table 1 also shows that teachers had an overall positive attitude towards listening and listening instruction.

Table 1. Descriptive statistics at the student and class level

Student-level characteristics				
		Mean (SD)	**Min**	**Max**
1.	Support at home	57.00 (14.00)	16.67	97.67
2.	Learning material	7.59 (0.90)	1	8
3.	Working memory	0.00 (1.01)	-5.81	1.62
4.	Vocabulary	-0.01 (0.67)	-3.01	1.87
5.	Intrinsic motivation	77.07 (13.25)	19.05	100
6.	Extrinsic motivation	57.71 (17.62)	16.67	100
7.	Listening strategy use	68.19 (14.41)	19.70	100
Class-level characteristics				
		Mean (SD)	**Min**	**Max**
8.	Herfindahl index	0.20 (0.19)	-0.19	0.50
9.	Class size	19.97 (5.07)	2	30
10.	Class resources	3.37 (0.96)	1	5
11.	Experience	14.43 (10.16)	1	38
12.	Time instruction	4.11 (1.28)	1	7
13.	Time evaluation	3.01 (1.12)	1	8
14.	Strategy instruction	64.67 (10.82)	41.67	100
15.	Teacher attitude	77.18 (11.49)	50	100
16.	Expectations	74.67 (15.33)	33.33	100
17.	Listening skills	0.06 (0.03)	-2.96	2.97

Table 2 gives an overview of the Pearson's bivariate correlations between the different variables, showing that the values of the correlates between the explanatory variables were rather low. As such, the assumption of no perfect multicollinearity was respected and no variables had to be removed. Table 2 shows that teachers with positive attitudes related to listening and listening instruction spend more time on evaluating listening, but do not integrate listening instruction more frequently in their lessons. Further, the results revealed that non-native Dutch speaking students score significantly lower on the verbal working memory test and the vocabulary test. Finally, no significant relationship was found between gender and vocabulary, or gender and working memory.

Table 2. Correlates at the student and class level

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Student-level characteristics																									
1. Gender	1																								
2. Home language	.01	1																							
3. Learning disability	-.11[b]	-.09[b]	1																						
4. Repeating a class	-.05	.17[b]	.06[a]	1																					
5. Educational level of the mother	0	-.26[b]	.02	-.10[b]	1																				
6. Educational level of the father	-.03	-.11[b]	-.04	-.11[b]	.56[b]	1																			
7. Educational support at home	.19[b]	.01	-.03	.01	.10[b]	.08[a]	1																		
8. Learning material at home	.06[b]	-.05	-.08[a]	0	.07[a]	.11[b]	.23[b]	1																	
9. Working memory	.06	-.08[a]	-.15[b]	-.09[b]	.17[b]	.12[b]	.06	.09[a]	1																
10. Vocabulary	-.04	-.20[b]	-.07[a]	-.02	.16[b]	.16[b]	.08[b]	.13[b]	.18[b]	1															
11. Intrinsic motivation	.12[b]	0	-.10[b]	-.06	0	.02	.35[b]	.16[b]	.05	.03	1														
12. Extrinsic motivation	-.10[b]	.12[b]	-.07[b]	.06	-.13[b]	-.10[b]	.13[b]	.02	-.05	-.10[b]	.28[b]	1													
13. Listening strategy	.03	.06	-.08[b]	-.02	-.03	-.03	.35[b]	.16[b]	.06	.02	.61[b]	.33[b]	1												
Class-level characteristics																									
14. Herfindahl index	.04	.52[b]	-.09[b]	.11[b]	-.22[b]	-.12[b]	.03	-.09[b]	-.03	-.11[b]	.02	.11[b]	.08[b]	1											
15. Class size	.02	-.02	-.07[a]	-.01	-.01	.04	.05	.08[b]	.02	.01	.01	0	0	-.03	1										
16. Class resources	-.06	-.11[b]	.03	-.10[b]	.12[b]	.06	.01	.12[b]	.04	.02	.09[b]	-.03	.01	-.13[b]	0	1									
17. Teacher gender	-.04	0	.05	.02	.04	.03	.02	.02	.03	.06	.01	.04	0	0	0	.11[b]	1								
18. Experience	.06	-.11[b]	-01	-.01	.05	.08[a]	-.03	.02	.09[b]	.08[a]	-.03	-.05	-.03	-.15[b]	.19[b]	.04	.21[b]	1							
19. Certification	.06[b]	.14	-.03	.01	-.05	-.04	.03	-.06	.01	-.06	.02	.02	.04	.10[b]	-.16[b]	.03	-.15[b]	.05	1						
20. Time listening	0	.02	-.05	.01	.04	.10[b]	.05	-.01	.05	.09[b]	-.01	-.01	-.05	.05	.02	-.16[b]	.05	-.03	-.13[b]	1					
21. Time evaluation	.02	.08[a]	-.05	.03	-.07[a]	-.04	.01	-.07[a]	-.03	.01	.02	.03	0	.27[a]	-.25[a]	.06	-.04	-.08[a]	.06	.41[b]	1				
22. Strategy instruction	-.02	.05	-.03	.01	-.01	.04	.03	-.03	.01	0	.07[a]	-.03	.02	.15[b]	.19[b]	.10[b]	.01	.15[b]	-.01	.26[b]	.27[b]	1			
23. Teacher attitude	-.10[b]	0	-.04	.07[a]	.02	.04	.03	.09[b]	.01	.09[b]	-.12[b]	.04	-.15[b]	.04	.19[b]	-.03	.11[b]	.12[b]	-.04	.12[b]	-.04	.26[b]	1		
24. Expectations	-.14[b]	.01	-.08[a]	.13[b]	.04	.09[b]	.03	.10[b]	.03	.05	-.09[b]	.04	-.15[b]	-.05	.01	.04	-.14[b]	-.24[b]	.25[b]	.12[b]	.03	.29[b]	-.14[b]	1	
25. Listening comprehension	-.03	-.17[b]	-.10[b]	-.16[b]	.26[b]	.26[b]	.13[b]	.13[b]	.27[b]	.23[b]	.03	-.12[b]	-.03	-.20[b]	.03	.09[b]	-.03	.10[b]	-.02	.07[a]	-.06	-.01	.11[b]	.13[b]	1

[a] Correlation is significant at the 0.05 level (2-tailed) [b] Correlation is significant at the 0.01 level (2-tailed)

5.2. Multilevel modeling

Model 0. The first step in the analysis was to explore whether multilevel modeling was required over a single level analysis to explain differences in elementary school students' listening skills. This model is referred to as the null model because it contains not one explanatory variable. The random part of the null model indicated that the variances at the class (σ^2_{u0} = .025, χ2 = 9.924, *df* = 1, *p* = .002) and the student level (σ^2_{e0} = .285, χ2 = 453.333, df = 1, *p* = .000) were significantly different from zero, justifying the application of multilevel modeling. The results indicate that the within-class differences are much larger than the between-class differences. The estimates σ^2_{u0} = .025 and σ^2_{e0} = .285 yield an intraclass correlation coefficient (ICC = σ^2_{u0} / (σ^2_{e0} + σ^2_{u0})) of .919, which indicates that 91.9% of the variance in students' listening skills is attributed to differences between individual students within classes, whereas 8.1% of the variance is due to differences between classes. The intercept of 0.058 should be interpreted as the overall mean of the score on the listening comprehension test of all students in all of the classes.

Model 1. In a following step, student control variables were added to the null model. For gender, being a girl was selected as the reference category; for home language, "Dutch-speaking" was chosen; whereas for the educational level of the parents, a parent with a higher education degree was selected as the reference category. The results showed that almost all included background variables, i.e., home language, class retention, learning disorder, educational level of the mother and the father, and educational support at home, with the exception of gender and educational resources, significantly contributed to the variation in listening skills. For example, Dutch-speaking students (ß = .149, χ^2 = 8.848, *df* = 1, *p* = .003) reported significantly higher listening skills than students who do not speak the instruction language as their first language. Further, the estimates revealed that students with a learning disorder (ß = -.139, χ^2 = 8.233, *df* = 1, *p* = .004) and students who already repeated a class (ß = -.229, χ^2 = 11.462, *df* = 1, *p* = .001) scored significantly lower for listening skills. Further, students with a mother with a degree of higher education scored significantly higher than students with a mother with a certificate of lower secondary education (ß = -.191, χ^2 = 6.348, *df* = 1, *p* = .012) or elementary education (ß = -.424, χ^2 = 10.052, *df* = 1, *p* = .002). Accordingly, students with a father with a higher education degree scored significantly higher than students with a father with a degree of higher secondary education (ß = -.137, χ^2 = 11.105, *df* = 1, *p* = .001), lower secondary education (ß = -.234, χ^2 = 11.423, *df* = 1, *p* = .001) or elementary education (ß = -.351, χ^2 = 7.330, *df* = 1, *p* = .007). Besides, students from whom the parents provide more educational support at home (ß = .054, χ^2 = 7.123, *df* = 1, *p* = .008) scored significantly higher for listening skills. The positive regression coefficient indicates that the more educational support students get at home, the higher they score on the listening test. Finally, girls did not perform better than boys for listening (*p* = .087) and educational resources at home (*p* = .197) were not related to students'

listening skills. Based on a comparison of the deviance, Model 1 fits the data significantly better than Model 0 (χ^2 = 222.674*, df* = 6, $p < .001$).

Model 2. After controlling for student background characteristics, different cognitive, linguistic, and motivational student variables, i.e., working memory, vocabulary, cognitive listening strategy use, and intrinsic and extrinsic listening motivation were added to the model. These variables were centered around their grand mean. The fixed effect of working memory (ß = .099, χ^2 = 33.126*, df* = 1, $p < .001$) was significant, showing that students with a higher working memory level tended to achieve higher scores for listening skills. Further, the effect of vocabulary on students' listening skills was significant (ß = .106, χ^2 = 16.619*, df* = 1, $p < .001$). The positive slopes indicate that for every increase with one unit, the score on the listening scale increased by 0.107. In addition, students with a higher extrinsic listening motivation scored significantly lower on listening skills (ß = -.003, χ^2 = 6.066*, df* = 1, p = .014) than students who were less extrinsically motivated. Further, students' intrinsic listening motivation (p = .577) and listening strategy use (p = .150) were not related to listening skills. After removing the non-significant parameters, Model 2b fitted the data significantly better than Model 1 (χ^2 = 39.456*, df* = 3, $p < .001$).

Model 3. In the next step of model specification, the teacher-level background variables, i.e., gender, years of experience, certification, language diversity, class size, and class resources were added to the model. Language diversity, years of experience, and class size were centered around their grand mean. The fixed main effect of language diversity in the class was highly significant (ß = -.392, χ^2 = 16.821, *df* = 1, $p < .001$), showing that students in more language diverse classes scored significantly lower for listening skills. Surprisingly, the influence of home language at student level became non-significant. Further, none of the other teacher-level background characteristics was significantly related to students' listening skills. After removing the non-significant parameters, Model 3b fitted the data significantly better than Model 2b (χ^2 = 14.667, *df* = 0, $p < .001$).

Model 4. In Model 4, teacher-level attitudes and practices were added. All variables were centered around their grand mean. Teacher attitudes, teacher expectations, and teacher practice variables, i.e., instructional time, evaluation time, and strategy instruction made no significant contribution to the model. As such, Model 4 did not fit the data significantly better than Model 3b. Model 3b was perceived as the final model concerning factors related to differences in sixth-grade students' listening skills in the Lsch. A summary of all model estimates can be found in Table 3, together with the standardized regression coefficients.

Table 3. Summary of the model estimates (dependent variable: score on listening comprehension test)

	Model 0 Null model	**Model 1a Student background model**	**Model 1b**	**Model 2a Student process model**	**Model 2b**	**Model 3a Teacher background (final model)**	**Model 3b**	**Model 4 Teacher process model**
Parameter	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)	B (SE)
Fixed part								
Intercept (Cons.)	.058 (.030)	.291 (.036)[b]	.263 (.031)[b]	.217 (.030)[b]	.218 (.030)[b]	.220 (.031)[b]	.201 (.028)[b]	.201 (.028)[b]
Student-level								
Gender (ref: girls)		-.061 (.035)						
Home language (ref: Dutch)		.141 (.050)[b]	.149 (.050)[b]	.093 (.046)[a]	.096 (.046)[a]			
Learning disorder		-.143 (.049)[b]	-.139 (.048)[b]	-.091 (.045)[a]	-.091 (.044)[a]	-.086 (.047)[a]	-.100 (.047)[a]	-.086 (.047)[a]
Educational level of the mother								
• Elementary		-.421 (.133)[b]	-.424 (.134)[b]	-.358 (.130)[b]	-.358 (.130)[b]	-.376 (.129)[b]	-.305 (.116)[b]	-.411 (.192)[b]
• Lower secondary		-.185 (.076)[a]	-.191 (.076)[a]					
• Higher secondary								
• Higher education (ref)								
Educational level of the father								
• Elementary		-.351 (.132)[a]	-.351 (.130)[a]	-.310 (.126[a]	-.315 (.125)[a]	-.289 (.125)[a]	-.351 (.132)[a]	-.331 (.117)[a]
• Lower secondary		-.226 (.069)[b]	-.234 (.069)[b]	-.183 (.068)[b]	-.185 (.068)[b]	-.192 (.067)[b]	-.226 (.069)[b]	-.194 (.066)[b]
• Higher secondary		-.134 (.041)[b]	-.137 (.041)[b]	-.123 (.040)[b]	-.125 (.040)[b]	-.140 (.039)[b]	-.134 (.041)[b]	-.132 (.039)[b]
• Higher education (ref)								
Educational support at home		.054 (.021)[a]	.054 (.020)[a]	.058 (.021)[a]	.053 (.020)[a]	.051 (.019)[a]	.056 (.019)[a]	.056 (.019)[a]
Educational resources at home		.192 (.149)						
Repeating a class		-.234 (.068)[b]	-.229 (.068)[b]	-.203 (.066)[b]	-.202 (.066)[b]	-.199 (.065)[b]	-.197 (.064)[b]	-.203 (.064)[b]
Working memory				.102 (.018)[b]	.099 (.018)[b]	.104 (.017)[b]	.101 (.017)[b]	.097 (.017)[b]
Vocabulary				.107 (.026)[b]	.106 (.026)[b]	.112 (.026)[b]	.108 (.025)[b]	.106 (.025)[b]
Motivation								
• Intrinsic				-.001 (.002)				
• Extrinsic				-.003 (.001) [a]	-.003 (.001)[a]	-.003 (.001)[a]	-.003 (.001)[a]	-.003 (.001)[a]
Listening strategy use				-.002 (.002)				

Table 3 (continued)

	Model 0	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b	Model 4
Class-level								
Class resources						.017 (.019)		
Class size						-.002 (.004)		
Language diversity (Herfindahl)						-.390 (.110)[b]	-.392 (.096)[b]	-.392 (.096)[b]
Teacher experience						.003 (.002)		
Teacher qualifications								
• Bachelor elementary (ref)								
• Bachelor secondary						.113 (.084)		
• Master without Teacher						-.119 (.168)		
• Master with Teacher						-.072 (.109)		
Teacher gender						-.079 (.041)		
Teacher attitude								.003 (.002)
Teacher expectations								.032 (.025)
Cognitive strategy instruction								-.021 (.027)
Time instructing listening								
• Multiple times a day								-.203 (.157)
• Every day								-.040 (.135)
• Multiple times a week								-.037 (.137)
• Once a week								.000 (.144)
• 2 or 3 times a month								.109 (.153)
• Once a month								.008 (.165)
• Multiple times a year								.000 (.000)
• Once a year (ref)								
Time evaluating listening								
• Multiple times a day								-.116 (.164)
• Every day								-.235 (.171)
• Multiple times a week								-.285 (.171)
• Once a week								-.297 (.193)
• 2 or 3 times a month								.000 (.000)
• Once a month								-.268 (.247)
• Multiple times a year								-.395 (.222)
• Once a year (ref)								

Table 3 (continued)

	Model 0	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b	Model 4
Random part								
Class level (between)	.025 (.008)	.008 (.005)	.008 (.005)	.005 (.004)	.006 (.004)	.000 (.003)	.004 (.004)	.000 (.003)
Student level (within)	.285 (.013)	.260 (.013)	.261 (.013)	.245 (.012)	.245 (.012)	.243 (.012)	.241 (.012)	.240 (.012)
Deviance (-2LL)	1596.579	1369.441	1373.905	1328.462	1334.449	1286.353	1319.782	1304.285
$\chi 2$		227.138	222.674	45.443	39.456	48.096	14.667	15.497
df		8	6	5	2	6	1	5
p		.000	.000	.000	.000	.001	.000	.153
Reference model	Single-level	Model 0	Model 0	Model 1b	Model 1b	Model 2b	Model 2b	Model 3b

[a] Correlation is significant at the 0.05 level (2-tailed); [b] Correlation is significant at the 0.01 level (2-tailed)

Note: β* are standardized coefficients; values in parentheses are standard errors.

Finally, the proportion of explained variance at the student and class level was obtained by calculating the squared multiple correlation coefficient R. Therefore, both R_1 at the student level (i.e., the proportional reduction of error for predicting an individual outcome with $R_1 = 1 - ((\sigma^2_{e0} + \sigma^2_{u0})_{\text{conditional model}}/(\sigma^2_{e0} + \sigma^2_{u0})_{\text{unconditional model}})$) and R_2 at the class level (i.e., the proportional reduction of error for predicting a group mean with $R_2 = 1 - (((\sigma^2_{e0} / \tilde{n} + \sigma^2_{u0})_{\text{conditional model}}/((\sigma^2_{e0} / \tilde{n}) + \sigma^2_{u0})_{\text{unconditional model}})$) were calculated (Van Laere et al., 2014) with "ñ" referring to the average cluster size ($\tilde{n}$ = 18.53) (Vaughn et al., 2019). Table 4 shows the proportion of explained variance at both the student (R_1) and the class level (R_2) for the estimated models. The proportion of explained variance that can be attributed to the integrated variables in the two models is identified by ΔR^2_1 and ΔR^2_2. In comparison with Model 1 (student background model), the proportion of explained variance in Model 2b increases by 5.8% at the student level and 7.7% at the class level, which is due to the integration of working memory, vocabulary, and extrinsic listening motivation. Moreover, the addition of language diversity in the class results in an extra proportion of variance of 1.9% and 5.0% at the student and class level respectively. The final model explains 57.9% of the variance in students' listening skills.

Table 4. Proportion of explained variance at the student and class level

	Model 1b	Model 2b	Model 3b
R_1 (proportion of variance explained at student level)	.132	.190	.210
ΔR^2_1		.058	.019
R_2 (proportion of variance explained at class level)	.453	.530	.579
ΔR^2_2		.077	.050

6. DISCUSSION

Despite the importance of listening skills in the elementary class, research explaining differences in students' listening skills in the Lsch is scarce (Andringa et al., 2012; McKendry & Murphy, 2010). Besides, existing listening studies are mainly focusing on isolated student-level correlates and do not take into account the multilayered structure in which they are embedded. This study adds to the literature by investigating the degree to which different student- and class-level factors are related to sixth-grade students' listening skills. To allow the nested structure of students within classes, a multilevel analysis was used. The results justified the application of multilevel modeling as a significant amount of variance in listening skills was attributed at the class level. However, the impact of student-level variables on listening skills is apparently of more significance than characteristics at the class level. These findings are in line with previous literacy studies showing that most of the original variance is due to differences between students within classes (e.g., De Smedt et al., 2016).

Our first interest was to explore the relationship between different cognitive, linguistic, and motivational student-level characteristics and listening skills. Except for listening strategy use and intrinsic listening motivation, the included student process characteristics turned out to be significantly related to students' listening skills. In line with previous studies (Kim, 2016; Molloy, 1997; Tighe et al., 2015), a higher verbal working memory capacity is positively correlated with students' listening skills. This makes sense, as the working memory is the mental workspace where students simultaneously store and process incoming information to establish coherence in text comprehension (Kim, 2016). Further, the present findings confirm previous studies showing that vocabulary knowledge is a foundational linguistic skill for listening comprehension (e.g., Kim, 2016; Marx & Roick, 2012; Wolfgramm et al., 2016). Unsurprisingly, a higher proportion of unknown words may cause lexical gaps, which can disturb the listening process. Considering motivational characteristics, the results showed that students with a higher extrinsic listening motivation performed significantly lower for listening skills than students who are less extrinsically motivated. This can be explained by the short-term effect of extrinsic motivation, as external pressure may be helpful for reaching short-term goals, but may be less effective in long-term learning processes (Nagel, 2011), such as listening. Surprisingly, there was no relationship between intrinsic listening motivation and listening skills. These patterns among the motivation orientations merits further exploration in follow-up research. Likewise, the use of cognitive listening strategies was not significantly associated with listening skills. In this way, the scales may not have been sensitive enough to detect differences in strategy use and motivation, or may have been subject to responder or recall bias as self-reported questionnaires often deal with over- or underreporting. Therefore, in future research, the questionnaires could be accompanied by interviews or think-alouds, which could provide more in-depth information about student motivation or strategy use.

Confirming the findings of earlier studies, non-native speaking students experienced a greater challenge in listening comprehension in the Lsch than their native speaking classmates (Marx & Roick, 2012; Wolfgramm et al., 2016). The descriptive results of this study also revealed that non-native speaking students had a disadvantage in the area of verbal working memory and had a more limited vocabulary knowledge in the Lsch, which could have a mediating effect on their performance in listening comprehension. However, it was remarkable that the relationship between home language and students' listening skills became non-significant when language diversity at the class level was included in the model. These results indicate that the differences in listening skills between native and non-native speaking students can be partially explained by the language diversity in the class. Earlier effectiveness studies did already confirm the important role of class composition on student outcomes. In Flanders, non-native Dutch-speaking students are not equally distributed across classes (Denessen, Driessen, & Bakker, 2010; Van Laere et al., 2014). As such, the amount of native speaking classmates can compensate for home language as non-native speaking students can be privileged of language opportunities due to the

higher level of language proficiency of their native speaking peers. Further, following general language research, the results revealed that students with higher educated parents and students who are receiving more educational support at home significantly outperformed their peers for listening skills. It is possible that higher educated parents are capable to create a richer listening environment for children or foster better attitudes towards learning and listening (Chiu, 2010). It was remarkable that in this study the fathers' educational level was of more significance at predicting listening achievement than the mothers' educational level.

The second research hypothesis investigated the relationship between class-level characteristics and students' listening skills. The results revealed that no significant relationship with teacher practices, such as listening strategy instruction or the amount of instructional listening time could be found. However, these findings may be influenced through the limited amount of listening instruction in elementary classes. In line with earlier research (Beall et al., 2008; Imhof, 2008), teachers in the present study indicated they barely spent time on instructing listening skills, making it difficult to investigate the influence of those instructional practices. There are different explanations for the lack of instruction: teachers may feel they are not properly trained to instruct listening skills, have not enough material available to teach and evaluate listening skills, or assume that children automatically know how to listen effectively in the class (Fogelsong, 2016; Siegel, 2014). This study raised the awareness for the need to provide further research opportunities on listening instructional practices in elementary schools. For example, the effect of listening strategy instruction should be investigated apart and longitudinally in intervention research, which include educational packages to improve teaching. For optimal results, teachers need to know the theories and principles behind listening strategy instruction. As such, future research could explore if explicit listening strategy instruction, where teachers model the listening strategies and provide opportunities for students to practice, may result in listening comprehension improvement. Further, the results revealed a positive attitude of teachers towards listening and listening instruction, but teacher attitudes were not significantly associated with students' listening skills. A possible explanation is that teacher attitudes will influence listening more indirectly through their association with instructional practices, but listening instruction is rather scarce in the upper elementary school context.

Finally, some limitations could be pointed out in this correlational study. A first important limitation concerns the use of a cross-sectional design in this study with data collected at a single point in time. Future research using a longitudinal design with multiple time points is still needed in order to clarify the stableness of listening skills over time. Second, in this study, only quantitative methods were used. Qualitative research could complement this work by providing more insight about the cognitive and motivational processes of the students and the practice of the teachers. For example, class observations could give a clearer view of instructional practices of the teacher in listening education. Third, future research should meaningfully elaborate on the relationships between the different factors. For example, the

relationship between student motivation and listening skills may be mediated by students' strategy use. A combination of both multilevel and structural equation modeling techniques could be considered to explore such causal paths. A final limitation is that we did not include school-level factors in this multilevel study. The low number of teachers per school made it impossible to investigate a three-level model and to measure additional variance at the school level. Future research could include school-level factors such as school size, turnover or dropout rates, and the availability of a high-quality educational school policy on the development of students' language skills. In spite of these limitations, the contribution of this study is clear. This study adds to the limited listening research by identifying important factors related to upper elementary school students' listening skills. The present study showed that verbal working memory, vocabulary knowledge, extrinsic listening motivation, different indicators of family background, and the language diversity in the class contribute to variation in students' listening achievement in the Lsch. It is of interest to identify and understand these key factors so that teachers can focus on policies that could improve listening education quality. For example, focusing on enlarging vocabulary knowledge promises to have a positive impact on listening comprehension. In particular, students with a learning disorder or non-native speaking students might benefit from vocabulary training. Further, training students' ability to store oral input in the working memory should improve their performance in listening comprehension tasks. Finally, the importance of listening skills is not in proportion with the instructional time spent on listening. Therefore, it could be a critical first step to train pre-service teachers in teaching listening skills and to provide them with sufficient coursework and resources for instruction. In conclusion, this study contributes to the unexplored field of listening research by adding solid empirical evidence of what constitutes effective listening.

ACKNOWLEDGMENTS

This work was supported by the Research Foundation Flanders (FWO) (under grant numbers 11B2417N). The data were collected and processed according to the General Data Protection Regulation (GDPR). This study was approved by the ethical commission of Ghent University.

REFERENCES

Acat, M.B., Demiral, H., & Kaya, M. F. (2016). Measuring listening comprehension skills of 5th grade school students with the help of web-based system. *International Journal of Instruction, 9*(1), 211-224. https://doi.org/10.12973/iji.2016.9116a

Adelmann, K. (2012). The art of listening in an educational perspective. Listening reception in the mother tongue. *Education Inquiry, 3*(4), 513-534. https://doi.org/10.3402/edui.v3i4.22051

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and Writing: An Interdisciplinary Journal, 19*(9), 933–958. doi:10.1007/s11145-006-9024-z

Akbari, R., & Dadvand, B., (2011). Does formal education make a difference? A comparison of pedagogical thought units of BA vs. MA teachers. *Modern Language Journal, 95*(1), 44-60. https://doi.org/10.1111/j.1540-4781.2010.01142.x

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition, 6*(2), 81-96. https://doi.org/10.1017/S1366728903001056

Alonzo, C. N., Yeomans-Maldonado, G., Murphy, K. A., & Bevens, B. (2016). Predicting second grade listening comprehension using prekindergarten measures. *Topics in Language Disorders, 36*(4), 312-333. doi:10.1097/TLD.0000000000000102

Anderson, J. R. (2000). *Cognitive psychology and its implication (5th ed.).* New York, NY: Worth Publishers.

Anderson, D. (2012). Hierarchical Linear Modelling (HLM): An introduction to key concepts within cross-sectional and growth modelling frameworks (Technical Report No. 1308). Eugene, OR: Behavioural Research and Teaching, University of Oregon.

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: an individual differences approach. *Language Learning, 62*(2), 49-78. https://doi.org/10.1111/j.1467-9922.2012.00706.x

Baleghizadeh, S., & Rahimi, A. H. (2011). The relationship among listening performance, metacognitive strategy use and motivation from a self-determination theory perspective. *Theory and Practice in Language Studies, 1*(1), 61-67. https://doi.org/10.4304/tpls.1.1.61-67

Beall, M. L., Gill-Rosier, J., Tate, J., & Matten, A. (2008). State of the context: listening in education. *International Journal of Listening, 22*(2), 123-132. https://doi.org/10.1080/10904010802174826

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation, 18*(6), 1-13.
Retrieved from https://scholarworks.umass.edu/pare/vol18/iss1/6

Belfi, B., Haelermans, C., & de Fraine, B. (2016). The long-term differential achievement effects of school socioeconomic composition in primary education: A propensity score matching approach. *British Journal of Educational Psychology, 86*(4), 501-525. https://doi.org/10.1111/bjep.12120

Berne, J. (2004) Listening comprehension strategies: A review of the literature. *Foreign Language Annals 37*(4), 521-533. https://doi.org/10.1111/j.1944-9720.2004.tb02419.x

Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–student interaction: Differences in relation to student prior attainment and elementary vs. secondary schools. *Learning and Instruction, 21*(6), 715-730. https://doi.org/10.1016/j.learninstruc.2011.04.001

Bourdeaud'hui, H., Aesaert, K., & van Braak, J. [submitted]. Development and validation of an audiovisual listening test to identify differences in primary school students' listening skills. *Language Assessment Quarterly.*

Brown, G. (2008). Selective listening. *System, 36*, 10-21. https://doi.org/10.1016/j.system.2007.11.002

Brownell, J. (2012). *Listening: Attitudes, principles, and skills (5th ed.).* New York, NY: Pearson.

Brysbaert, M., Keuleers, E., Mandera; P., & Stevens, M. (2014). *Woordenkennis van Nederlanders en Vlamingen anno 2013: Resultaten van het Groot Nationaal Onderzoek Taal* [Vocabulary knowledge of Dutch and Flemish speakers in 2013: Results of the 'Groot Nationaal Onderzoek Taal.']. Ghent, Belgium: Academia Press.

Chang, A. C., & Read, J. (2013). Investigating the effects of multiple-choice listening test items in the oral versus written mode on listeners' performance and perceptions. *System, 41*(3), 575-586. https://doi.org/10.1016/j.system.2013.06.001

Chiu, M. M. (2010). Inequality, family, school, and mathematics achievement. *Social Forces, 88,* 1645-1676. doi:10.1353/sof.2010.0019

Clark, K. A., Scruggs, T. L., & Szydlowski, S. L. (1999). *Improving listening skills in the classroom.* (Doctoral dissertation, Saint Xavier University).

Cohen, S., & Wolvin, A. (2011). Listening to stories: An initial assessment of student listening characteristics. *Listening Education Online Journal, 3*(2), 16-25. Retrieved from http:// www.listen.org/Resources/Documents/LE%202011-2.pdf

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment. Research & Evaluation, 10*, 1-9. Retrieved March 5, 2013 from http://pareonline.net/getvn.asp?v = 10&n = 7

Creemers, B., & Kyriakides, L. (2010). School factors explaining achievement on cognitive and affective outcomes: Establishing a dynamic model of educational effectiveness. *Scandinavian Journal of Educational Research*, *54*(3), 263-294. https://doi.org/ 10.1080/00313831003764529

Croninger, R. G., Rice, J. K., Rathbun, A., & Nishio, M. (2007). Teacher qualifications and early learning: effects of certification, degree, and experience on first-grade student achievement. *Economics of education review, 26*(3), 312-324. https://doi.org/10.1016/j.econedurev.2005.05.008

De Smedt, F., Van Keer, H., & Merchie, E. (2016). Student, teacher and class-level correlates of Flemish late elementary school children's writing performance. *Reading and Writing: An Interdisciplinary Journal, 29*(5)*,* 833-868. https://doi.org/10.1007/s11145-015-9590-z

Denessen, E., Driessen, G., & Bakker, J. (2010). School and classroom diversity effects on cognitive and non-cognitive student outcomes. *Journal of Education Research, 4*(2), 1-13. Retrieved from https://www.researchgate.net/publication/235006628_School_and_classroom_diversity_effects_on_cognitive_and_non-cognitive_student_outcomes

Denis, J. Dierick, S., Janssen, R., & Aesaert, K. (2019). *Peiling Nederlands in het basisonderwijs—Brochure. [Dutch survey in primary education - Brochure].* Leuven, The Netherlands: KU Leuven en UAntwerpen, Steunpunt voor Toetsontwikkeling en Peilingen.

Driessen, G. (2003). Family and child characteristics, child-rearing factors, and cognitive competence of young children. *Early Child Development and Care, 173*(2-3), 323-339. https://doi.org/10.1080/0300443031000071879

Dronkers, J. & van der Velden, R. (2013). Positive but also negative effects of ethnic diversity in schools on educational performance? An empirical test using PISA data. *Tijdschrift voor Onderwijsrecht en Onderwijsbeleid*, *52*, 483–499. https://doi.org/10.1007/978-94-007-6119-3_4

Field, J. (2004). An insight into listeners' problems: Too much bottom-up or too much top-down? *System, 32,* 363-377. https://doi.org/10.1016/j.system.2004.05.002

Field, J. (2010). Listening in the language classroom. *ELT Journal, 64*(3), 331-333. https://doi.org/10.1093/elt/ccq026

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology, 27*(4), 935–951. https://doi.org/10.1348/026151008x397189

Fogelsong, D. F. (2016). *Promoting a pedagogy for listening instruction: Primary grade Teachers perceptions of teaching listening through interactive read alouds* (Doctoral dissertation, Virginia Tech).

Fredriksson, P., Ockert, B., & Oosterbeek, H. (2013). Long-term effects of class size. *Quarterly Journal of Economics, 128,* 249–285. https://doi.org/10.1093/qje/qjs048

Gliem, J. & Gliem, R. (2003). Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales. Paper presented at the 2003 Midwest Research to Practice Conference in Adult, Continuing and Community Education, the Ohio State University, Columbus, Ohio. Retrieved from https://scholarworks.iupui.edu/bitstream/handle/1805/344/Gliem+&+Gliem.pdf?sequence = 1.

Goh, C. (2008). Metacognitive instruction for second language listening development: Theory, practice and research implications. *RELC Journal, 39*(2), 188-213. https://doi.org/10.1177/0033688208092184

Graham, S. (2017) Research into practice: listening strategies in an instructed classroom setting. *Language Teaching, 50*(1), 107-119. https://doi.org/10.1017/S0261444816000306

Graham, S., D., Santos, R. & Vanderplank. (2011). Exploring the relationship between listening development and strategy use*. Language Teaching Research, 15*(4), 435–56. https://doi.org/10.1177/1362168811412026

Green, R. (2017). *Designing listening tests: a practical approach*. Basingstoke, UK: Palgrave Macmillan.

Hagtvet, B. E. (2003). Listening comprehension and reading comprehension in poor decoders: Evidence for the importance of syntactic and semantic skills as well as phonological kills. *Reading and Writing: An Interdisciplinary Journal, 16*(6), 505-539. https://doi.org/10.1023/A:1025521722900

Harputlu, L., & Ceylan, E. (2014). The effects of motivation and metacognitive strategy use on EFL listening proficiency. *Procedia-Social and Behavioral Sciences, 158*, 124-131.

https://doi.org/10.1016/j.sbspro.2014.12.056
Hassel, S., & Ridout, N. (2017). An investigation of first-year students' and lecturers' expectations of university education. *Frontiers in Psychology, 8,* 2218. https://doi.org/10.3389/fpsyg.2017.02218
Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech-Language Pathology, 16*(3), 199–207. https://doi.org/10.3109/17549507.2014.904441
Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariances structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118
Huddleston, A. P. (2014). Achievement at whose expense? A literature review of test-based grade retention policies in U.S. schools. *Education Policy Analysis Archives, 22*(18), 1-34. https://doi.org/10.14507/epaa.v22n18.2014
Imhof, M. (2008). What have you listened to in school today? *International Journal of Listening, 22*(1), 1-12. https://doi.org/10.1080/10904010701802121
Jalongo, M. R. (2008). *Learning to listen, listening to learn: Building essential skills in young children.* Washington, DC: National Association for the Education of Young Children.
Jeffries, S., & Everatt, J. (2004). Working memory: its role in dyslexia and other specific learning difficulties. *Dyslexia, 10*(3), 196-214. https://doi.org/10.1002/dys.278
Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149. https://doi.org/10.1037/0033-295X.99.1.122
Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627-633. https://doi.org/10.3758/BRM.42.3.627
Kim, Y.-S. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology, 141*, 101-120. https://doi.org/10.1016/j.jecp.2015.08.003
Klauda, S. L. (2009). The role of parents in adolescents' reading motivation and activity. *Educational Psychology Review, 21*(4), 325-363. https://doi.org/10.1007/s10648-009-9112-0
Kyriakides, L., Creemers, B., & Charalambous, E. (2019). Searching for differential teacher and school effectiveness in terms of student socioeconomic status and gender: implications for promoting equity. *School Effectiveness and School Improvement, 30*(3), 286-308. https://doi.org/10.1080/09243453.2018.1511603
Lau, K.-L. (2017). Strategy use, listening problems, and motivation of high- and low-proficiency Chinese listeners. *The Journal of Educational Research, 110*(5), 503-514. https://doi.org/10.1080/00220671.2015.1134421
Lehto, J. E., & Antilla, M. (2003). Listening comprehension in elementary level grades two, four and six. *Scandinavian Journal of Educational Research, 47*, 133–143. https://doi.org/10.1080/00313830308615
Lin, S. W., Liu, Y., Chen, S. F., Wang, J. R., & Kao, H. L. (2015). Development of a computer-based measure of listening comprehension of science talk. *International Journal of Science and Mathematics Education, 13*(6), 1469-1486. https://doi.org/10.1007/s10763-014- 9559-4
Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: examining where these differences lie. *Journal of Research in Reading, 32*(2), 199-214. https://doi.org/10.1111/j.1467-9817.2008.01389.x
Marx, A., & Roick, T. (2012). Prädiktoren des Hörverstehens bei Jugendlichen deutscher und Jugendlichen nichtdeutscher Herkunftssprache [Predictors of listening comprehension in native and non-native German speaking adolescents]. *Zeitschrift für Pädagogische Psychologie, 26,* 121–134. https://doi.org/10.1024/1010-0652/a000067
McInnes, A., Humphries, T., Hogg-Johson, S., & Tannock, R. Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *Journal of Abnormal Child Psychology, 31*(4), 427-443. https://doi.org/10.1023/A:1023895602957
McKendry, M. G., & Murphy, V. A. (2010). A comparative study of listening comprehension measures in English as an additional language and native English-speaking primary school children. *Evaluation & Research in Education, 24*(1), 17-40. https://doi.org/10.1080/09500790.2010.531702

Molloy, P. J. (1997). The role of individual differences in working memory in reading and listening comprehension in intermediate grade students. Doctoral dissertation. Available from *Dissertation Abstracts International Section A: Humanities and Social Sciences, 58*(6-A), 2072.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/pirls2016/international-results/

Nagel, M. C. (2011). Motivation in the middle: Looking back, moving forward. *Australian Journal of Middle Schooling, 11*(2). Retrieved from http://eds.a.ebscohost.com/

Oduolowu, E. O., & Oluwakemi, A. E. (2014). Effect of storytelling on listening skills of primary one pupil in Ibadan north local government area of Oyo state, Nigeria. *International Journal of Humanities and Social Science, 4*(9), 100–107.
Retrieved from http://www.ijhssnet.com/journals/Vol_4_No_9_July_2014/10.pdf

Osada, N. (2004). Listening comprehension research: A brief review of the past thirty years. *Dialogue, 3*(1), 53-66. Retrieved from http://www.talk-waseda.net/dialogue/no03_2004/2004dialogue03_k4.pdf

Palardy, G., & Rumberger, R. W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis, 30*(2), 11-140. https://doi.org/10.3102/0162373708317680

Rasbash, J., Browne, W. J., & Charlton, C. (2018). MLwiN version 3.02. Centre for Multilevel Modeling, University of Bristol (2018).

Raudenbush, S. W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and data analysis methods (2nd ed.).* Thousand Oaks, CA: Sage Publications, Inc.

Roch, M., Florit, E., & Levorato, M. C. (2012). The advantage of reading over listening text comprehension in Down syndrome: what is the role of verbal memory? *Journal of Research in Developmental Disabilities, 33*, 890-899. https://doi.org/10.1016/j.ridd.2011.11.002.

Ryan, R. M., & Deci, E. L. (2002). Overview of self-determination theory: An organismic dialectical perspective. In E. L. Deci, & R. M. Ryan (Eds.), *Handbook of self-determination research (pp. 3–33).* Rochester, NY: University of Rochester Press.

Siegel, J. (2014). Second language learners' perceptions of listening strategy instruction. *Innovation in Language Learning and Teaching, 7*(1), 1–18.
https://doi.org/10.1080/17501229.17502011.17653110.

Staehr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31*(4), 577–607.
https://doi.org/10.1017/S0272263109990039

Stutz, F., Schaffner, E., & Schiefele, U. (2017). Measurement invariance and validity of a brief questionnaire on reading motivation in elementary students. *Journal of Research in Reading, 40*(4), 439-461. https://doi.org/10.1111/1467-9817.12085

Suvorov, R. (2015). Interacting with visuals in listening tests: An eye-tracking study. In V. Berry (Ed.), *ARAGs research reports online* (Report #AR-A/2015/1). Retrieved from https://lib.dr.iastate.edu/cgi/viewcontent.cgi?referer = https://www.google.com/&httpsredir = 1&article = 4306&context = etd

Thys, S., & Van Houtte, M. (2016). Ethnic composition of the elementary school and educational choice: does the culture of teacher expectations matter? *Teaching and Teacher Education, 59*, 383-391. https://doi.org/000383813500034

Tighe, E. L., Spencer, M., & Schatschneider, C. (2015). Investigating predictors of listening comprehension in third, seventh, and tenth grade students: A dominance analysis approach. *Reading Psychology, 36*(8), 700-740. https://doi.org/10.1080/02702711.2014.963270

Ulu, M. (2016). A structural equation model to explain the effect of fluent reading, literal comprehension and inferential comprehension levels of elementary school 4th grade students on success in problem solving. *Education and Science, 41*(186), 93-117. https://doi.org/10.15390/EB.2016.6303

Vandecandelaere, M., Vansteelandt, S., De Fraine, B., & Van Damme, J. (2016). The effects of early grade retention: Effect modification by prior achievement and age. *Journal of School Psychology*, *54*, 77-93. https://doi.org/10.1016/j.jsp.2015.10.004

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3-25. https://doi.org/10.1017/S0267190504000017

Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive awareness and proficiency in listening. *Applied Linguistics, 26*(1), 70-89. https://doi.org/10.1093/applin/amh039

Vandergrift, L. (2007). Recent development in second language listening comprehension research. In Graeme Porte (Ed.), *Language Teaching: Surveys and Studies,* 291-310. Montreal, Canada: Cambridge University Press.

Van den Bergh, L., Denessen, E., Hornstra, L., Voeten, M., & Holland, R. W. (2010). The implicit prejudiced attitudes of teachers: relations to teacher expectations and the ethnic achievement gap. *American Educational Research, 47*(2)*,* 497-527. https://doi.org/10.3102/0002831209353594

Van Laere, E., Aesaert, K., & van Braak, J. (2014). The role of students' home language in science achievement: A multilevel approach. *International Journal of Science Education, 36*(16), 2772-2794. https://doi.org/10.1080/09500693.2014.936327

Wolfgramm, C., Suter, N., & Göksel, E. (2016). Examining the role of concentration, vocabulary and self-concept in listening and reading comprehension. *International Journal of Listening, 30*(1-2), 25-46. https://doi.org/10.1080/10904018.2015.1065746

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806-838. https://doi.org/10.1177/0011000006288127

APPENDIX A: LISTENING TEST FRAMEWORK

Table A1. Test framework with components, skills and example items (translated from Dutch)

Component	**Subskill**	*Example items*
1.Literal	1.1. Defining the literal meaning of a word or a word group.	*Example item for skill 1.1** *What is the task of an alert dog?* A. Warning his owner. B. Curing his owner. C. Helping his owner. D. Protecting his owner.
	1.2. Remembering facts that have been explicitly mentioned in the auditory text.	*Example item for skill 1.2** *What does Sammy do when Emily's sugar level is too low?* A. Sammy starts barking. B. Sammy presses the alarm button. C. Sammy starts wagging. D. Sammy warns Emily's parents.
	1.3. Identifying detailed information. • persons and objects • numbers • place and time	
2.Inferential	2.1. Deriving the implicit meaning of a word or a word group. 2.2. Linking different sentences or text parts. • cause and effect • reason and explanation • comparison and contrast • means and ends 2.3. Identifying the global content or purpose of the text.	*Example item for skill 2.2** *Why is Sammy such a special dog?* A. He smells better than other dogs. B. He is the first dog that was trained by Laura. C. He is the first dog that can help children with diabetes. D. He is smarter than other dogs.

* These are example items from an informative text about a diabetes-alert dog.

APPENDIX B: VALIDATION WORKING MEMORY AND VOCABULARY TEST

Working memory test

To psychometrically validate the working memory test, all 72 items ran through five phases of analysis, i.e., 1) classical item analysis, 2) dimensionality, 3) model-data fit, 4) local independence, and 5) monotonicity. First, results of the classical item analyses showed all items of the working memory test discriminate well between a student who does well and who does poorly on the test. Second, results of confirmatory factor analysis indicated generally good fit of the single-factor model, i.e. unidimensionality (GFI: .99994; RMSR: .006). Eight items were deleted due to a factor loading below .300. Third, results from the non-linear factor analysis indicated that the 2PLM provided the best fit: SRMSR (1PLM: .070; 2PLM: .062) and MADaQ3 (1PLM: .047; 2PLM: .047). The 2PLM is preferable over the 1PLM, as the AIC decreases a lot (35513-35429=84) between the 1PLM and the 2PLM. Further, item 2 (.38; .22), item 12 (.57; .27), item 16 (.20) and item 17 (.35) had one or more Q3-values higher than .2, indicating that they were interrelated with one or more other items. Finally, no items violated the assumption of monotonicity. The remaining 60 items comprised the working memory test. Cronbach's alpha was .88, indicating good internal consistency of the scale.

Vocabulary test

To psychometrically validate the vocabulary test, all 54 items ran through five phases of analysis, i.e., 1) classical item analysis, 2) dimensionality, 3) model-data fit, 4) local independence, and 5) monotonicity. First, results of the classical item analyses showed the item-total correlation of 25 items was located outside the critical range of .15 or the p-value was not between .30 or .95. As these items cannot sufficiently discriminate between students or were too difficult or too easy, they were removed for further analysis, resulting in 29 items. Second, results of confirmatory factor analysis indicated generally good fit of the single-factor model, i.e. unidimensionality (GFI: .99978; RMSR: .00736). Nine items were deleted due to a factor loading below .30. Third, results from the non-linear factor analysis indicated that the 2PLM provided the best fit: SRMSR (1PLM: .046; 2PLM: .041) and MADaQ3 (1PLM: .042; 2PLM: .039). The 2PLM is preferable over the 1PLM, as the AIC decreases a lot (19257-19175=82) between the 1PLM and the 2PLM. Finally, no items violated the assumption of local independency or monotonicity. The remaining 20 items comprised the vocabulary test. Cronbach's alpha was .71, indicating relatively good internal consistency of the scale.

APPENDIX C: SCALE VALIDATION

Different steps were conducted in order to examine the psychometric properties of these scales. First, using SPSS Statistics 22, it was checked whether any of the items violated the assumption of normal distribution, i.e., if the kurtosis was between -2 and 2 and the skewness value was between -1 and 1. Second, in order to check the quality of our instruments, exploratory (EFA) and confirmatory factor analysis (CFA) is used. For the validation of the developed student and parent questionnaire, the original student sample (n = 974) was divided into four equally sized subsamples. This allowed us to conduct different EFA replication analysis on subsample 1, 2 and 3, and a CFA on subsample 4. Maximum likelihood extraction with direct oblimin (oblique) rotation was used to find the number of latent variables that belonged to each construct (Beavers et al., 2013). Factor extraction was based on the eigenvalue-greater-than-1 rule (Costello & Osborne, 2005) and examination of the scree plots supported the factor solution (Beavers et al., 2013). Further, a CFA was used to assess whether the proposed factor structure of the EFAs fits the data well. The fit indices: Standardized Root Mean square Residual (SRMR), Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker Lewis Index (TLI) were mentioned in order to check the hypothesized structure (Worthington & Whittaker, 2006). For the RMSEA and SRMR lower values indicate good model fit, although for the CFI and TFI, values greater than .90 are desirable (Hu & Bentler, 1999). For the validation of the adapted motivation scale, only a CFA was conducted on the total sample. For the teachers, the scale validation was conducted on the participating teachers in this study (n = 70) and an additional sample of teachers who only filled in the teacher questionnaire (n = 50). As the total sample size for the validation of the teacher questionnaire was only 120, a single CFA analysis was conducted on the total sample. Finally, the reliability of the developed scales was checked. When using Likert-type scales, it is desirable to calculate and report Cronbach's alpha coefficient for internal consistency reliability for the scales, with scores higher than .7 indicating acceptable and scores higher than .8 indicating good internal consistency of the factors (Gliem & Gliem, 2003).

APPENDIX D: STUDENT-LEVEL SCALES

Table D1. Factor loadings for the listening strategies items of the students (translated from Dutch)

Six-point scale ranging from 1 (totally disagree) to 6 (totally agree)		EFA	EFA	EFA	CFA
		SS1 (*n*=302)	**SS2 (*n*= 310)**	**SS3 (*n*=319)**	**SS4 (*n*=285)**
1.	Before I start listening, I have a plan in my head about how I am going to listen.	.62	.58	.64	.51
2.	Before I start listening, I listen to the title of the text to predict what will happen.	.69	.58	.53	.55
3.	As I listen, I try to figure out what the purpose of the text is.	.54	.57	.54	.46
4.	As I listen, I subdivide the text into different text parts.	.57	.49	.51	.48
5.	As I listen, I try to imagine and visualize what is told.	.52	.50	.48	.46
6.	As I listen, I try to figure out the meaning of unfamiliar words.	.67	.60	.62	.65
7.	As I listen, I search for the main topic of the listening text.	.62	.57	.72	.58
8.	As I listen, I try to concentrate me as good as possible on the listening text.	.61	.43	.55	.61
9.	After listening, I check if what I heard is true and makes sense.	.57	.52	.61	.59
10.	After listening, I try to summarize what has been told.	.65	.56	.61	.59
11.	After listening, I wonder if I understood everything what has been told.	.66	.57	.58	.70

Model Fit: SS4 TLI: .93, CFI: .95, RMSEA: .06, SRMR: .05, Cronbach's alpha: .84

Table D2. Factor loadings for the listening motivation items (translated from Dutch)

Six-point scale ranging from 1 (totally disagree) to 6 (totally agree)	**CFA**	
	Intrinsic motivation	Extrinsic motivation
1. I listen because I am interested in what has been told.	.78	
2. I listen because that's the way to learn more about interesting topics.	.71	
3. I listen because that's the way I can learn something new.	.74	
4. I listen because the stories that are told are thrilling me.	.73	
5. I listen because it is fun.	.73	
6. I listen because I am curious about what the teacher has to say.	.60	
7. I listen because I want to perform well in listening.		.61
8. I listen because that's the way I can discover new and difficult words.		.70
9. I listen because I want to outperform others in the class.		.26
10. I listen because it is important to be one of the best listeners in the class.		.28
11. I listen because other people think it is good for me to listen well.		.28
12. I listen because I can talk to my parents and friends about what was told.		.52

Model Fit[a]: TLI: .91, CFI: .94, RMSEA: .08, SRMR: .06

[a]TLI (Tucker-Lewis Index), CFI (Comparative Fit Indices), RMSEA (Root Mean Square Error of Approximation), SRMR (Standardized Root Mean Square Residual)

Table D3. Factor loadings for the educational support at home items (translated from Dutch)

Six-point scale ranging from 1 (never) to 6 (always)		**EFA**	**EFA**	**EFA**	**CFA**
		SS1 (*n* = 321)	**SS2 (*n* = 323)**	**SS3 (*n* = 315)**	**SS4 (*n* = 319)**
1.	I talk to my parents about what's happening at school.	.56	.48	.54	.38
2.	My parents help me with the computer.	.53	.50	.56	.45
3.	My parents buy different books for me.	.58	.59	.58	.59
4.	My parents read the newspaper or books with me.	.54	.54	.53	.49
5.	My parents take me to the museum, concert or theater.	.64	.54	.54	.55
6.	My parents take me to the library.	.55	.49	.44	.52
7.	I talk to my parents about the news.	.48	.56	.54	.54
8.	My parents help me with my homework.	.38	.29	.35	.34
9.	My parents listen to the radio with me.	.41	.46	.38	.26

Model Fit: SS4 TLI: .92, CFI: .94, RMSEA: .05, SRMR: .05, Cronbach's alpha: .74

APPENDIX E: CLASS-LEVEL SCALES

Table E1. Factor loadings for the instruction in listening strategies items (translated from Dutch)

Six-point rating scale ranging from 1 (never) to 6 (always)	**CFA (n=120)**
1. Remembering facts or details from the listening text.	.35
2. Deriving difficult words from the listening text.	.42
3. Identifying the main topic from the listening text.	.60
4. Linking the information to the prior knowledge of the text.	.65
5. Identifying the purpose of the text or the text type.	.60
6. Summarizing the listening text.	.54
7. Judging the truthfulness or accurateness of the listening text.	.55
8. Judging the reliability of the listening text.	.56

Model Fit: TLI: .95, CFI: .91, RMSEA: .08, SRMR: .06, Cronbach's alpha: .79

Table E2. Factor loadings for the teacher attitudes items (translated from Dutch)

Six-point rating scale ranging from 1 (totally disagree) to 6 (totally agree)	**CFA (n=120)**
1. I think teaching listening skills is important.	.73
2. I think it is important for students to acquire a good level of listening.	.55
3. I like teaching listening skills.	.73
4. I think spending time on listening education is a waste of time, because listening cannot be taught.	.05
5. I think spending time on listening education is as important as spending time on other language skills.	.77
6. I think evaluating listening skills is as important as evaluating other language skills.	.59
7. When I am running out of time, I will spend less time in listening education.	.10
8. I think spending time on listening education is a waste of time, because listening is an innate skill.	.13
9. I think students with good listening abilities will perform better at secondary and high school.	.49
10. I think students with good listening abilities are more likely to find a good job.	.43

Model Fit : TLI: .92, CFI: .95, RMSEA: .09, SRMR: .06, Cronbach's alpha: .76

APPENDIX F: SCALE OVERVIEW

Table F1. Description of the measured variables in the student and teacher questionnaire

Variable	Description	Item format	Nr. items
Student background			
✓ Student gender	Gender (S)	Binary	1 item
✓ Home language	Language most spoken with mother (S)	Nominal	1 item
✓ Disorder	Developmental disorders (P)	Binary	8 items
✓ Educational level of the parents	Highest educational level of the mother and the father (P)	Nominal	2 items
✓ Educational support at home	Parental assistance at home (S)	Likert type	9 items
✓ Learning material	Learning material (S)	Binary	9 items
✓ Repeating a class	If students already repeated a class (S)	Binary	1 item
Teacher background			
✓ Experience	Teacher experience in years (T)	Continuous	1 item
✓ Teacher certification	Type of certification of the teacher (T)	Nominal	1 item
✓ Teacher gender	Gender (T)	Binary	1 item
Class composition			
✓ Class resources	Material in the class (T)	Binary	6 items
✓ Class size	Number of students per class (T)	Continuous	1 item
✓ Herfindahl index	Language diversity (T)	Continuous	1 item
Student process			
✓ Working memory	Verbal working memory capacity (S)	Continuous	60 items
✓ Vocabulary	Receptive word recognition (S)	Continuous	20 items
✓ Intrinsic motivation	Intrinsic listening motivation (S)	Likert type	6 items
✓ Extrinsic motivation	Extrinsic listening motivation (S)	Likert type	6 items
✓ Listening strategy	Listening strategy use (S)	Likert type	11 items
Teacher process			
✓ Strategy instruction	Instruction in listening strategies (T)	Likert type	8 items
✓ Teach time listening	Time spent on instructing and evaluating listening skills (T)	Likert type	2 items
✓ Teacher attitudes	Attitudes towards listening (T)	Likert type	10 items
✓ Teacher expectations	Teacher expectations scale (T)	Likert type	3 items

S= student questionnaire, P= parent questionnaire, T= teacher questionnaire